Information Platform for adolescent education

# 1 市场定位

# 2 命名、注册

# 3 组网架构

建设原则：低成本、低投入、高可靠性、大容量带宽、两地多中心、大容量存储的

## 3.1 云平台

### 3.1.1 免费额度

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 云服务器  (cpu/mem/net/disk/free days) | 云数据库  (mem/disk/free days) | 域名  (month) | CDN  (free traffic) |  |
| 腾讯云 | 1C/1G/5Mbps/50GB/180 | 1G/50GB/180 | 12 | 10GBytes/mon |  |
|  | 2C/4G/5Mbps/50GB/96 | 2G/100GB/96 | 12 | 10GBytes/mon |  |
|  | 4C/8G/5Mbps/50GB/60 | 4G/200GB/45 | 12 | 10GBytes/mon |  |
| AWS |  |  |  |  |  |

### 3.1.2 收费额度

## 3.2 自建平台

成品服务器/自己DIY服务器 + 多线接入 + 多WAN路由器/低端交换机 + SATA阵列盒 + SATA硬盘/二手SATA硬盘

# 4 应用开发

# 5 数据爬取

## 5.1 基于Web Driver + Httpwatch + Html Parser的页面数据爬取

1、如何面对千变万化的页面结构，而采用更高效的手段获取；

2、如何保证数据的时效性、

## 5.2 电话咨询

# 6 大数据算法

## 6.1 客户画像

数据结构化

### 6.1.1 必要的一些数据

1、普通家长/学生/幼儿：孩子兴趣爱好、孩子年龄、空闲时间、父母年收入范围、居住地址

2、培训机构：课程类型（特长班、考级班）、课程名称、课程特色、任课老师、课程费用（年/月/次/课程是否可以顺延）

3、课程评价、评分；

### 6.1.2 匹配规则

根据向量阈值+权重，匹配最符合孩子兴趣范围、年龄、空闲时间、父母收入及距离远近的高质量培训机构、课程。

## 6.2 数据库

HANA or neo4j or NOSQL or MySQL